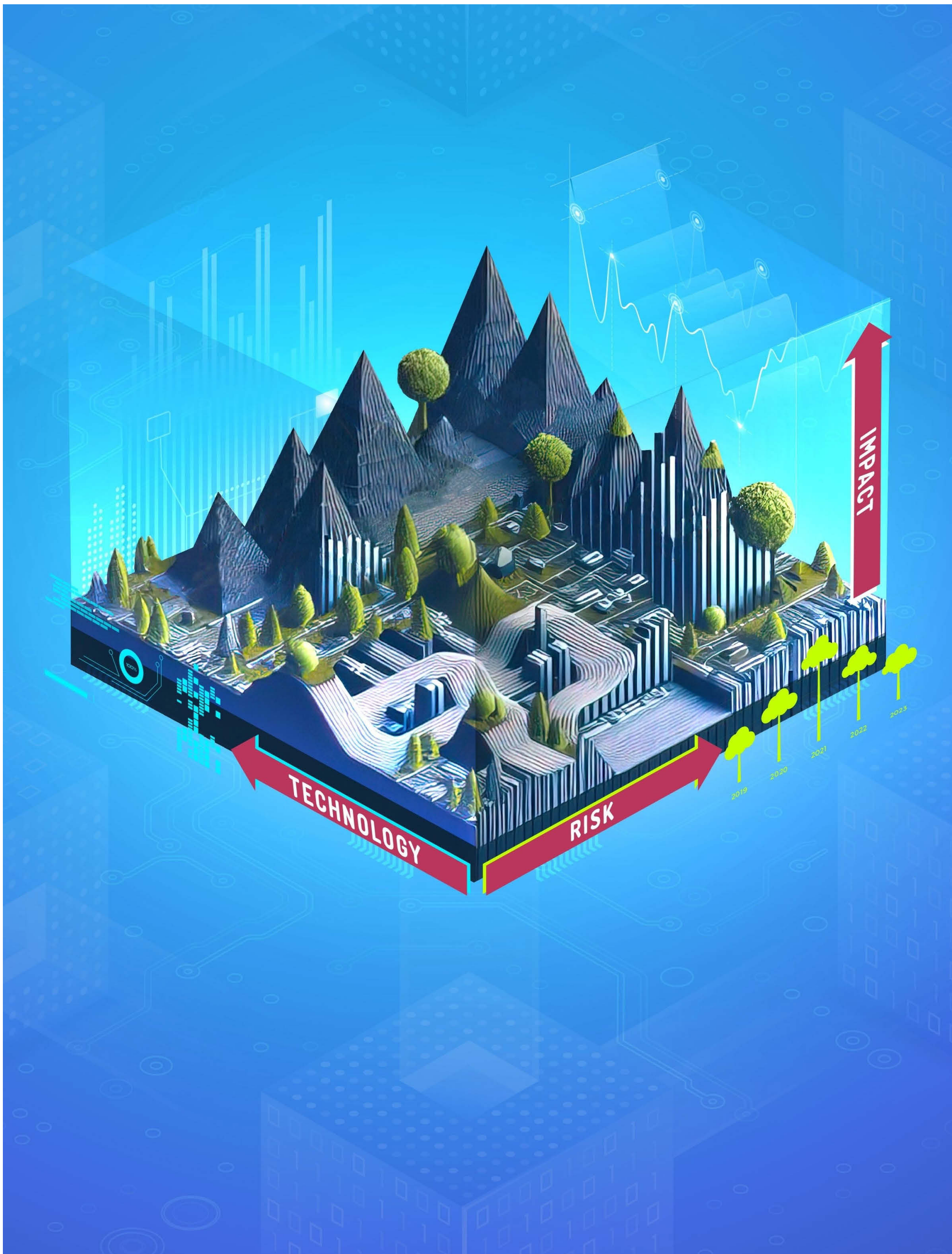
Lucerne University of Applied Sciences and Arts

AI and Impact Investing

Institute of Financial Services Zug IFZ
www.hslu.ch/ifz



FH Zentralschweiz

Contents

1. Introduction

With the rapid development of artificial intelligence systems, especially large language models (LLMs) and their ability to process natural language, various applications of this technology seem possible in almost any field. The financial sector, which has large amounts of information at its disposal, could benefit from this development by further simplifying the extraction and analysis of relevant data to streamline operations, enhance decision-making, and improve customer experiences. This also applies to sustainable investing, an area that is becoming increasingly important as investors seek greater transparency and accountability in the sustainability practices of companies they invest in.

The growing product area of sustainability-related investments, which account for about 52 percent of the total Swiss fund market as of the end of 2022 (Busch et al., 2023), underlines the increasing relevance of sustainability to the financial sector in Switzerland. This growth has generally been fuelled by corresponding customer demand and changing regulations. Additionally, products related to sustainability demonstrate a positive effect on the margins of financial institutions (Swiss Bankers Association, online) which is advantageous from the providers' perspective. To assess their sustainability strategies, an increasing number of investment funds distributed in Switzerland are adopting the Sustainable Development Goals (SDGs) as benchmarks (Mattman & Stüttgen, 2023).

While there is a growing popularity for sustainable investments, significant challenges persist in accurately evaluating the sustainability performance of funds and consequently the companies they invest in. The present report aims to explore the potential of applying LLMs to quantify the SDG exposure of companies, and thus aims to leverage the potential of natural language processing for assessing sustainable investments. The choice of the SDGs as the basis for measuring sustainable impact is due to their popularity. This is underlined by a market survey of the Global Impact Investing Network (2023) which reveals that over three-quarters of impact investors use the SDGs as a conceptual framework to guide their impact strategy as of the end of 2022.

Specifically, the present report develops a prototype to create a company ranking of the 17 SDGs for each of the companies underlying two exemplary investment funds. For this purpose, company-specific information is extracted from public websites and subsequently analysed using LLMs from OpenAI to assess its relevance to each of the SDGs.

It is evident that the developed prototype does not address the underlying issues of SDG targets. However, it could facilitate the review of SDG alignment for specific companies and investment funds, potentially validating corresponding SDG assessments and providing transparency for investors. The report starts with a review of disciplines from the field of Artificial Intelligence in Chapter 2, laying the groundwork for describing LLMs in Chapter 3. Subsequently, Chapter 4 highlights key aspects of impact investing and the SDGs, and Chapter 5 details the architecture of the developed prototype. In Chapter 6, the derived empirical results from the prototype are discussed and corresponding potentials and limitations are addressed in Chapter 7. Finally, Chapter 8 concludes the report and provides an outlook.[1]

We would like to thank the research partners of the "IFZ FinTech Program", who have supported this condensed study financially as well as in terms of content. These are e.foresight, Finnova, InventxLab, SIX, Swiss Bankers Prepaid Services, and Zürcher Kantonalbank.

[1] Note that some of the content presented is adapted from the reports *Quantum Computing and AI in Finance* by Ankenbrand, Rhyner, and Yilmaz (2023), and *GPT for Financial Advice* by Ankenbrand, Bieri, et al. (2023).

2. Artificial Intelligence

Artificial Intelligence (AI) refers to the capacity of machines to carry out tasks that typically demand human intelligence, including tasks like visual perception, speech recognition, decision-making, and language processing (Bildirici, Ozge Zeytin, 2023). In 1950, Alan Turing introduced the idea of testing whether machines can exhibit human-like intelligence through what became known as the "Turing Test". This test involves a human evaluator interacting with both a human and a machine via written communication, trying to discern which is which. Success in the Turing Test means that the machine's responses are indistinguishable from those of a human. For this, the machine must effectively communicate in natural language (Natural Language Processing), retain and utilise information (Knowledge Representation), deduce answers and make logical inferences (Automated Reasoning), and adapt to new situations while recognising and extrapolating patterns (Machine Learning). Though Turing's focus was on intellectual capabilities rather than physical appearance or actions, subsequent researchers have expanded this concept into what is termed the "total Turing test", which involves real-world interactions with objects and individuals. To pass such a test, a robot would require abilities such as recognition for perceiving the environment (Computer Vision) and manipulation of objects and mobility (Robotics). These six domains encapsulate most of AI (Russell & Norvig, 2016). While Robotics and Computer Vision fall under these domains, they do not directly align with the goals of this report and will not be discussed further. The remaining domains have a closer connection to the objectives and will be explored in more detail in the following paragraphs.

The objective of Machine Learning is to develop algorithms capable of learning independently, without the need for human input or guidance. Rather than directly programming the computer to tackle a task, Machine Learning seeks methods for the computer to create its own solutions based on provided examples and objectives (Grosan & Abraham, 2011). This approach is particularly beneficial for tasks where algorithms are unclear, but ample data is accessible. When leveraging historical data to discern regularities and patterns, Machine Learning can provide valuable predictive insights. This capability has a wide range of applications, e.g. in credit assessment, fraud detection, the prediction of stock market trends (Alpaydin, 2020), but also in language processing.

Natural Language Processing (NLP), which can be seen as a subdomain of Machine Learning, outlines the machine-controlled and artificial intelligence-based area of automatically analysing and representing human language to establish effective communication between computers and humans (Cambria & White, 2014). Classical tasks associated with NLP are text classification, question answering, sentiment analysis, machine translation, and speech recognition.

In NLP, Knowledge Representation can be operationalised in different ways depending on the techniques or models used. For example, information might be encoded in knowledge graphs or databases. Knowledge graphs allow structuring information regarding real-world objects or abstract concepts, describing relationships and semantic properties to specify connections and dependencies between objects (Ji, Pan, Cambria, Marttinen, & Yu, 2022). Knowledge Representation is often closely knit to Automated Reasoning technologies, which can be used to extract new information and conclusions from data. Among other techniques, a reasoning process might include discovering relationships, applying logic, and validating new information using existing knowledge (Chen, Jia, & Xiang, 2020). Hence, such simulated thinking processes might be used to solve many of the mentioned tasks associated with NLP.

In connection with the four domains of AI mentioned, the development of Large Language Models (LLMs) currently stands out in particular, as they are showing advanced capabilities in generating coherent and fluent text based on inputs provided by a user (Khan, Daud, Khan, Muhammad, & Haq, 2023). While LLMs do not directly engage in traditional knowledge representation and automated reasoning, they interact with these concepts by processing and generating language that can include structured information and reasoning. Although their ability to handle complex language tasks can sometimes mimic reasoning and knowledge application, it's important to note that their "understanding" is based on statistical correlations rather than logical reasoning or structured knowledge. However, given their advanced capabilities in language processing and semantic analysis, LLMs have been chosen as the foundational technology for the prototype developed in this report. The functioning of LLMs is further detailed in Chapter 3.

3. Large Language Models

Many AI solutions are currently based on approaches from the field of Large Language Models. This study considers the following definition of such models:

A Large Language Model (LLM) is a language model consisting of a neural network with billions of parameters trained on large amounts of unlabelled data by means of self-supervised learning and used, for example, to predict and generate text and other content (Sejnowski, 2023).

A "Generative Pre-Trained Transformer (GPT)" is a specific type of LLM characterised by its layered architecture, as depicted in Figure 1. A GPT consists of multiple transformer blocks, each containing a "Self-Attention" mechanism and a "Feed Forward Neural Network". The self-attention mechanism enables the model to assign importance to different words and phrases within the input, enhancing its understanding of their context and semantic relationships. Meanwhile, the feed forward neural network within each transformer block is designed to capture nonlinear relationships between the inputs and outputs. The output of the transformer block is processed by an output layer ("Next Word Prediction Head") that generates a probability distribution over the vocabulary to determine the most likely next words. Depending on the model's configuration, a word is selected from this distribution. This selected word is then appended to the input sequence, and the model iteratively repeats this process to generate text.

Given their objective of understanding intricate word relationships in natural language, GPTs are characterised by their vast parameter count, established through training on extensive datasets. The inherent complexity of GPT models renders them uninterpretable, as their input-output relationships lack straightforward explication (Basu, Varanasi, Shakerin, Arias, & Gupta, 2021).

In practical terms, current GPTs generate text word-by-word sequentially, leveraging probability distributions derived from the training data to yield coherent responses to prompts such as questions. Consequently, the probabilistic nature of GPTs renders their responses sensitive to nuanced variations in input wording or phrasing, resulting in potential disparities in output consistency (Bubeck et al., 2023).[1] Moreover, GPTs' probabilistic framework may produce inaccurate content, commonly referred to as "hallucination" (Manakul, Liusie, & Gales, 2023).

Due to their capacity for processing and analysing language effectively, GPTs can be leveraged to review (public) data and identify and extract relevant key topics. More specifically, the model outputs can be optimised to yield responses that are firmly rooted in the content of such data. This approach is known as Retrieval-Augmented Generation (RAG) and deals with problems related to data unknown to the model in the training phase. This capability could make them valuable tools for content analysis and data scrutiny, potentially also in the domain of impact investing, an area of finance whose potential and challenges are discussed in the following Chapter 4.

[1] Note that output consistency can be influenced using parameters like nucleus sampling to control the variability of responses.

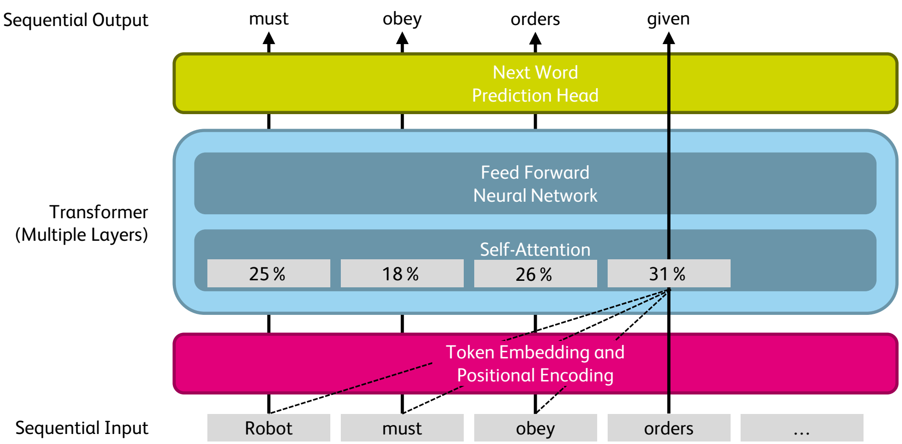


Figure 1: Simplified architecture of a Generative Pre-Trained Transformer (GPT) (source: based on Alammar (2019))

4. Impact Investing and SDGs

Apart from the general popularity of sustainable investment approaches, impact investing is growing as part of the professionalisation of the philanthropy field. In particular, it can be described as funding companies focused on creating a measurable impact by solving social and ecological problems. As it can be differentiated from other sustainable investment approaches, such as socially responsible investing (SRI) into public stocks and bonds, these strategies are also often complementary (Spiess-Knafl & Scheck, 2017). As of the end of 2022, impact investors allocated more than half of their assets under management (54 %) into energy, financial services, healthcare, microfinance, and food & agriculture according to a market study by the Global Impact Investing Network (2023).

Due to its focus on improving social and ecological environments, impact investing is closely connected to the SDGs which were agreed on by all 193 United Nations (UN) member states in September 2015 (Kharas, McArthur, & Ohno, 2019). Since the majority of the 17 SDGs target societal problems with only three SDGs focus directly on environmental issues (i.e., climate change, life underwater, and life on land), the framework particularly allows the categorisation of social issues. Nevertheless, depending on the interpretation and context, the various goals intersect both areas, i.e., ecology and society (Mattman & Stüttgen, 2023).

Depending on the perspective, the 17 SDGs can be grouped to enable comprehensive transformation in the societal and ecological sphere. Sachs et al. (2019) provide a framework of six SDG transformations to activate government, business, and civil society actions to solve problems targeted by the SDGs. In particular, the transformations combine the SDGs into the areas (1) *education, gender, and inequality*, (2) *health, well-being, and demography*, (3) *energy decarbonisation and sustainable industry*, (4) *sustainable food, land, water, and oceans*, (5) *sustainable cities and communities*, and (6) *digital revolution for sustainable development*.[1] The Sustainable Development Solutions Network (SDSN), an initiative launched by the United Nations, is promoting these transformations inclusively among the different stakeholders ranging from scientists and governments to the private sector (SDSN Association, online).

Further frameworks and instruments for operationalising the SDG goals for investors and stakeholders exist, such as the "Impact Investing Market Map" of the UN Principles for Responsible Investment (UN PRI). This initiative is, for example, targeting mainstream impact investing entities, including listed equity firms, medium and large businesses, as well as infrastructure projects. The mainstream focus should assure high liquidity and maturity of investments. Furthermore, the thematic and financial terms for each investment are determined individually, based on essential certifications, initiatives, and benchmarks of company revenues in MSCI, FTSE, or Bloomberg indexes, alongside PRI data (Principles for Responsible Investment, 2018).

As sustainability jargon is sometimes vague and a company's entire sustainability performance is difficult to estimate (Van Zanten & Van Tulder, 2018), researchers are testing the consistency of different types of sustainability ratings under frameworks such as SDG and ESG (environmental, social, and governance). Nevertheless, the various ratings and frameworks are not always comparable. For example, in contrast to the impact-focused SDGs scores, ESG ratings primarily evaluate whether a company's financial performance could be impacted by ESG factors (Amel-Zadeh, 2017). For this reason, the SDGs are considered in the present study as a framework for assessing companies' commitment to sustainability.

While the growth of impact investing reflects the general popularity of sustainable investment solutions, challenges remain in accurately assessing companies' sustainability performance. However, aligned with the SDGs, impact investing seems to offer tangible frameworks for targeted action and collaborative efforts across sectors. In this context, (responsible) AI methodologies might enable the realisation of SDGs (Ametepey, Aigbavboa, Thwala, & Addy, 2024). As a practical pilot application based on LLMs, the following Chapter 5 presents a prototype that aims to evaluate public information on companies of two exemplary investment funds for their SDG commitment in order to improve transparency in this area.

[1] A detailed description of all 17 SDGs can be found on the UN website (see here).

5. Prototype

In general, the main functionality of the prototype is to gather public information and use it to generate a ranking of a company's exposure to the SDGs. For the current case study, data of the underlying companies of two exemplary private equity funds has been collected. While one fund is categorised as a "Carbon Fund" by the fund provider, the other is labelled a "Growth Fund".[1] In this context, the prototype can be used to assess if companies within funds vary in their SDG exposure.

In streamlining the processing and analysis of data, the presented prototype relies on the initial version of OpenAI's Assistant API. This API supports the review of company information, extraction of SDG-related data, and the subsequent creation of SDG rankings. Supplying the necessary files to the Assistant API simplifies the implementation of RAG applications. However, it is worth noting that the inability to fine-tune the retrieval strategy is a trade-off for ease of use, where the Assistant automatically selects between performing a vector search and incorporating the file content directly into the prompt, depending on the length of the documents. The GPT-4 Turbo model by OpenAI has been used to enable the virtual SDG expert.

The prototype's design and task flow are illustrated in Figure 2, displaying its four main components to generate SDG rankings. Firstly, the underlying companies of two private equity funds were identified and the domains of their public websites, which provide the data source for creating the desired ranking, were collected. Secondly, a web scraping module gathered the public information of the websites as HTML files and extracted the text components. After applying data-preparation measures to the saved text files, the company information was fed to the GPT Assistant in the third step. For example, the individual text files of each company were combined into one larger text file, including the page names as headers, to comply with the maximum of 20 files for each session of the assistant. The GPT Assistant, acting as an SDG expert, was asked to review the text files of each company, indicating which information is deemed relevant and to finally return a SDG ranking for each entity. For this step, multiple prompting approaches were tested to obtain the desired ranking based on the available, not purely SDG-specific information. The final prompt strategy includes the following elements:

1. A role assignment as an SDG expert
2. A question asking for a summary of the provided file to increase consistency across multiple iterations
3. A question asking for a list of relevant SDGs
4. A request to provide a ranking of the identified SDGs[2]

The derived ranking consists of the SDGs numbered from 17 to one. However, the prototype usually deems a lower number of SDGs important, creating missing values. The handling of missing values is explained in Chapter 6. Moreover, the assistant repeats the question-answering process for each company 60 times to achieve greater consistency in the derived results, measured by the convergence of the mean, median, and standard deviation. The question-answering is stopped after 60 iterations, i.e., once the control values have stabilised. The final SDG ranking is then derived from the median ranking of all iterations of the assistant in the fourth step.

[1] Note that the prototype methodology described in this chapter has no relation to the sustainability evaluation approach followed by the respective fund provider.

[2] Fourth prompt: "If you had to, how would you rank these (SDGs) against each other in terms of how prominent they seem to be in the company's self-presentation based on the information provided to you."

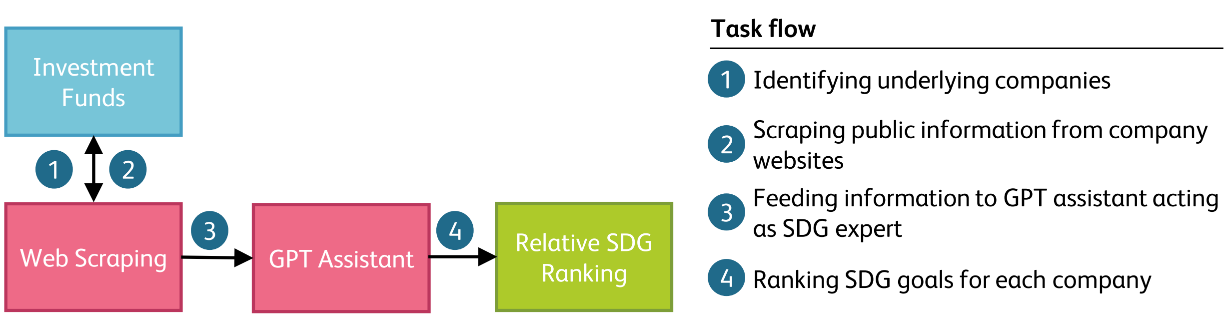


Figure 2: General architecture of the prototype and corresponding task flow

6. Empirical Results

The results discussed in this chapter are obtained using the prototype described in Chapter 5. Figure 3 illustrates the median score per SDG and company for the two exemplary investment funds over 60 iterations.[1] Note that the fund companies listed on the y-axes of the two graphs have been anonymised. In total, the growth fund comprises 15 companies, while the carbon fund contains 11 companies. Furthermore, in both graphs, the SDGs on the x-axes are sorted by the sum of the median values per SDG. The colour scale shows the importance of the SDGs by points, i.e., the most important SDG receives 17 points. Furthermore, the individual colour scale per plot indicates the maximum number of identified SDGs over all companies. Up to four SDGs were identified as relevant by the prototype for both the companies in the growth fund and the carbon fund. Consequently, as the prototype did not result in a ranking of all 17 SDGs in any iteration of the evaluation for a company, the unranked SDGs were assigned a score of zero points. Considering median values instead of taking an average over the iterations helps mitigate the influence of potential outliers and missing values. Such values expanded the distribution of recorded values per SDG, especially when using the default temperature setting of the LLM. Moreover, to exclude misleading median values from company rankings reporting an interquartile range greater than four, such values have also been replaced with zero.

[1] This ensures appropriate convergence of the median values for most SDGs for each company.

The left-hand graph of Figure 3 shows that the SDGs *9. Industry, Innovation, and Infrastructure*, *3. Good Health and Well-being*, and *12. Responsible Consumption and Production* are identified as the most relevant for the companies underlying the growth fund, according to the prototype's analysis of publicly available website data. The relatively high exposure to these SDGs may stem from their broad applicability and association with characteristics of high-growth companies, such as a focus on "innovation". Six of the 17 SDGs are not considered a priority for any of the companies.

In the right-hand graph of Figure 3, it appears that the companies underlying the reviewed carbon fund prioritise ecology-related SDGs comparably more. In detail, although the SDGs *12. Responsible Consumption and Production* and *9. Industry, Innovation, and Infrastructure* can also be found among the three most relevant, the goal *13. Climate Action* occupies the first position. Furthermore, with *7. Affordable and Clean Energy* a second ecology-related SDG is of comparably great relevance. These findings seem to correspond with the ecology focus of the investment fund. In comparison to the growth

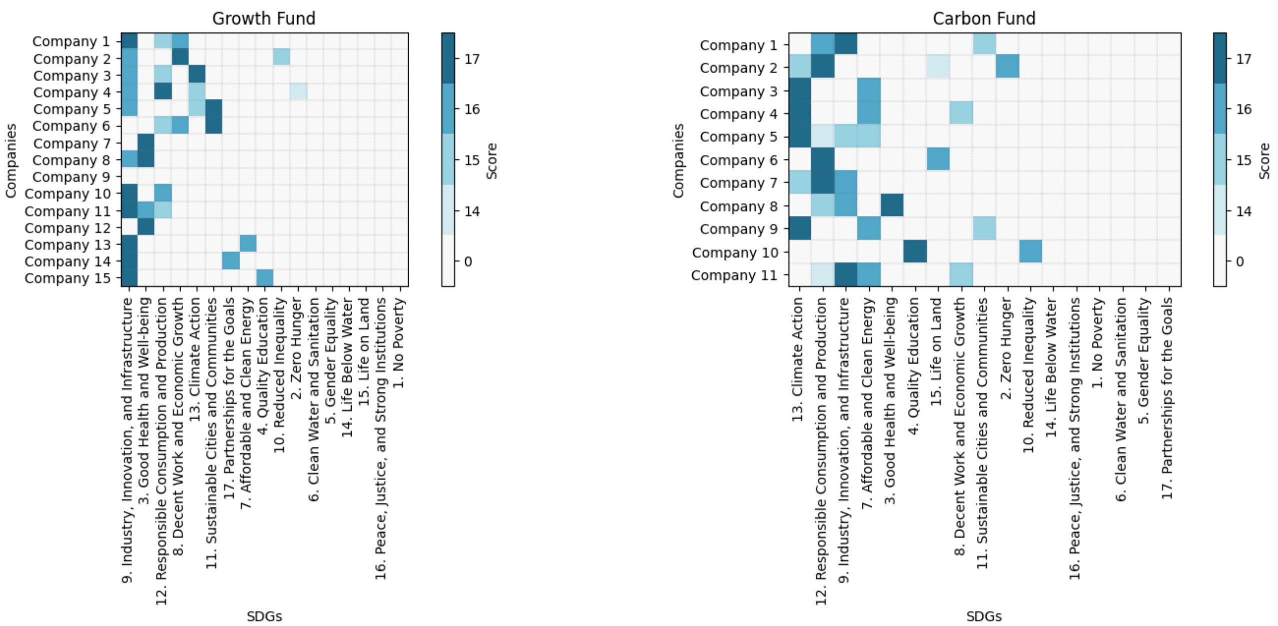


Figure 3: Median value per company and SDG for growth fund (left-hand graph) and carbon fund (right-hand graph)

fund, similar SDGs have not been considered a priority for any of the companies. These are *1. No Poverty*, *5. Gender Equality*, *6. Clean Water and Sanitation*, *14. Life Below Water*, and *16. Peace, Justice, and Strong Institutions*.

To compare the results of different GPT models, both GPT-3.5 Turbo and GPT-4 Turbo were used to assess the SDG commitment of the companies in the two funds. GPT-4 Turbo produced more consistent results, exhibiting less variability in the SDG assessment for each company, likely due to its broader general knowledge and reasoning skills compared to its predecessor (OpenAI, online). Additionally, the consistency of the results has been influenced by the "temperature" parameter to control the variability of responses for a specific input. A low value of 0.01 led to more deterministic, or less variable, results than the default value of one. Generally, lower temperature values produce more deterministic outcomes, indicating their impact on the variability of generated responses. Since the first version of the OpenAI Assistant API did not offer additional parameters for model specification, the scope for further customising the prototype was limited. This also includes enhanced control over retrieval strategies, which was only introduced in the second version of the API.

In summary, the developed prototype returns more general SDGs for companies that are components of the rather traditional growth fund. For the carbon fund companies, climate and energy specific SDGs seem to play a more important role. This highlights that the approach can differentiate between different degrees of exposure to the SDGs.

7. Potentials and Limitations

Although the prototype showcased the capabilities of LLMs in assessing sustainability impacts, it also highlighted areas for potential improvement and certain limitations. This chapter explores these aspects in detail, though the discussion is not exhaustive.

Prompt engineering is a newly emerging field dedicated to crafting and refining prompts to maximise the efficiency of language models in various applications and research domains. Prompts function as a type of programming that tailors the outputs and interactions with a language model to meet specific requirements. However, prompting can encounter certain pitfalls like ambiguity, bias reinforcement, overfitting, lack of context, ethical dilemmas, unintended side effects, and unrealistic reliance on model constraints (White et al., 2023). In the prototype presented, such concerns regarding context and ambiguity were addressed by incorporating warm-up prompts. These prompts assess the model's comprehension of the given documents. Subsequently, questions structured to emphasise the model's reliance solely on the provided data were utilised. Neglecting to do so often leads to ambiguous responses from the LLM, based on our observations.

In several studies on LLMs, it has been consistently demonstrated that larger models generally deliver superior performance. Nevertheless, as LLMs increase in size, the process of fine-tuning and managing all the parameters becomes prohibitively expensive and eventually becomes practically unfeasible (Ding et al., 2023). As an alternative, recent research has highlighted the efficacy of smaller language models in specific contexts. Despite their reduced size, these smaller models can still achieve competitive performance while offering advantages such as faster inference times and lower resource requirements (Sanh, Debut, Chaumond, & Wolf, 2020). By leveraging techniques such as knowledge distillation (Hinton, Vinyals, & Dean, 2015) and parameter pruning (Hassani & Khasahmadi, 2020), Small Language Models (SLM) can provide a more efficient and scalable solution for practical applications. In practice, it turns out that LLMs require significant computing power to provide users with sufficient performance. The power required for these operations is provided by graphics chips (GPU), which were originally created for the gaming industry. However, these are actually only available to a limited extent on the global market and are correspondingly expensive. SLMs, in contrast, can be operated on conventional computer chips (CPU) or smaller sized GPU-infrastructure. Commodity CPUs are generally available, are significantly cheaper, and, compared to GPUs, also consume significantly less energy ("Sustainable AI").

Another factor affecting the performance of language models is the amount and quality of data, as well as the model's complexity. The GPT-3.5 model is based on around 175 billion parameter - the latest GPT-4.0, according to the manufacturer, on around 100 trillion (OpenAI, online). Services like ChatGPT are based on a time-defined status (versioning) and cannot provide the latest knowledge beyond their cut-off date. However, many use cases in practice of the finance industry require company-specific data that is not publicly available. Such data can be added by the end-user to a limited extent using an upload function (e.g. ChatGPT) or the AI-application accesses the data available to it (e.g. Copilot in Microsoft Teams). But this raises the questions about data protection and compliance of such applications. The company-specific requirements regarding information security, as

Table 1: Operating models of language models applications

<table>
<tr><th>Criteria</th><th colspan="4">Characteristics</th></tr>
<tr><td>Use Case</td><td>Data Analytics</td><td>Chat Bots</td><td>Text Generation</td><td>...</td></tr>
<tr><td>Input Setup</td><td colspan="2">Continuous</td><td colspan="2">On Demand</td></tr>
<tr><td>Data Classification</td><td>Public</td><td>Internal</td><td>Confidential</td><td>Restricted</td></tr>
<tr><td>Transparency</td><td colspan="2">Black Box</td><td colspan="2">Understandable</td></tr>
<tr><td>Corporate Data Integration</td><td>RAG</td><td colspan="2">Mix</td><td>Fine-tuning</td></tr>
<tr><td>Deployment</td><td>SaaS</td><td>Public Cloud</td><td>Community Cloud</td><td>On Premise</td></tr>
<tr><td>Licence Type</td><td colspan="2">Commercial</td><td colspan="2">Open Source</td></tr>
<tr><td>Language Model</td><td colspan="2">LLM</td><td colspan="2">SLM</td></tr>
</table>

well as the legal requirements must be met, in particular the data protection laws (Federal Data Protection and Information Commissioner, 2023). In addition to the storage location and storage duration, attention should be paid to whether the data is used as a basis for training the language model or not. Questions also arise about intellectual property, respectively who owns the generated output of the AI - or how this can be used.

When using language models including company-specific data, it is recommended to optimise the data to reach optimal performance and output. To enhance the effectiveness of language models for specific use cases, there are two common approaches (Jeong, 2023): Fine-tuning and Retrieval-Augmented Generation (RAG), which can also be combined for improved results. RAG works by first retrieving relevant information from an external knowledge base. The retrieved information is combined with the original query as input into a language model. The generated answers are based on the model's pre-trained knowledge and on the use case-specific data. Fine-tuning adapts a pre-trained foundation model to a specific use case. The training of the model is continued on a smaller dataset, to adjust its parameters specifically for the task. In the prototype presented, only RAG was used. With RAG, the language model can always use up-to-date information, but the retrieving mechanism can be resource intensive. Fine-tuning usually needs less resources during deployment, but a training process is required (Ovadia, Brief, Mishaeli, & Elisha, 2023).

To summarise, there are different operating models depending on the specific application of language models and the existing framework conditions. The morphological box in Table 1 provides a synopsis of the most important criteria and associated options.

8. Conclusion and Outlook

The aim of this study was to analyse the potential of LLMs to assess the SDG commitment of companies in two selected funds. The core findings are summarised in the following statements and theses:

Artificial Intelligence is currently on everyone's agenda. According to Gartner's hype cycle, the valley of disappointment will soon come before a sustainable implementation and anchoring of AI in the banking industry arrives (Gartner, 2023). Chat and voice bots, for example, will become established in banking because significant efficiency advantages can be achieved in these processes. In practice, however, it also shows that the development and operation of specific solutions can be difficult and quite cost-intensive. It therefore seems all the more important that the framework conditions for the use of AI (trust, risk, and security) are prepared and that the necessary data classification and preparation are already tackled proactively. Last, but not least, there is the question of the extent to which customers accept AI-based solutions.

Impact investing is a relevant use case for an LLM-based prototype. Despite the rising popularity of impact investing, substantial challenges persist in accurately evaluating the sustainability performance of companies. The United Nations SDGs provide a structured framework for assessing a company's commitment to sustainability. In this context, the language and text processing capabilities of LLMs can be used to improve the analysis of sustainability data. More specifically, using the SDG framework as a benchmark, LLMs have the potential to analyse publicly available information about companies and assess their alignment with sustainability goals.

LLMs are able to carry out sustainability assessments. The prototype based on OpenAI's GPT-4 Turbo demonstrates that LLMs can generally analyse publicly available data for its relevance to the SDGs. The empirical analysis of companies from two funds, a growth fund and a carbon fund, shows that those from the latter tend to have a higher commitment to the ecologically oriented SDGs than the growth fund, which is therefore consistent with the overall focus of the fund. These findings justify further research in the field of LLMs and impact investing, and could ultimately lead to greater transparency in sustainability practices of companies and greater efficiency in corresponding assessments.

The application of LLMs is not trivial. Although standardised interfaces are available for the use of LLMs, the implementation of an application-specific prototype involves challenges and requires various data processing steps and model parameter choices. Additionally, the outputs of the models must be critically reviewed for quality and consistency, and appropriate adjustments made in cases of discrepancies. An understanding of LLMs and their limits is therefore necessary in order to create high-performance and efficient solutions.

There are some limitations of LLMs (and AI in general). Two factors that can lead to limitations in the performance of LLMs for certain use cases are data quality and the probabilistic nature of the systems. The effectiveness of LLMs in specific applications can be compromised if the data being analysed is of poor quality or irrelevant to the context required for the analysis, leading to unreliable outcomes. In the context of the prototype discussed, integrating additional data sources such as internal company information or other public datasets beyond just the company's website could further enhance result accuracy. Additionally, it is important to assess on a case-by-case basis whether the probabilistic behaviour of LLMs and their lack of interpretability are suitable for the specific application at hand.

For practical use, operations is key. To effectively utilise LLMs in operational business, it's essential to develop an operating model that aligns with both the technological and regulatory demands specific to the use case. Company-specific applications frequently necessitate the analysis of private information, which can present data protection challenges when using a publicly hosted LLM by a third party. In contrast, operating a proprietary LLM with billions of parameters demands substantial IT-infrastructure capabilities. A viable alternative to address these issues could be the adoption of Small Language Models, which can provide a more efficient solution for practical applications. From a business perspective, it is also important to maintain an awareness of the evolving regulations surrounding the field of AI.

Authors and Contributors

This condensed study was prepared in collaboration with the following individuals who contributed in the form of text, discussion, document reviews, and other forms of feedback (in alphabetical order).

Authors

Thomas Ankenbrand
Head Competence Center Investments
Lucerne University of Applied Sciences and Arts

Denis Bieri
Lecturer
Lucerne University of Applied Sciences and Arts

Levin Reichmuth
Research Associate
Lucerne University of Applied Sciences and Arts

Ege Yilmaz
Research Associate
Lucerne University of Applied Sciences and Arts

Guest Authors

Carla Caspar
Strategic Innovation Manager
Inventx Lab AG

Fabian Keller
Co-Lead Foresight & Innovation
Zürcher Kantonalbank

Attila Makra
Co-Lead Foresight & Innovation
Zürcher Kantonalbank

Urs Rhyner
Head of InventxLab
Inventx Lab AG

Contributors

Mike Eggenschwiler
Swiss Bankers Prepaid Services

Denise Gubser
e.foresight

Contact

For more information about this report, please contact us at:

Thomas Ankenbrand
Lucerne University of Applied Sciences and Arts
thomas.ankenbrand@hslu.ch

Fabian Keller
Zürcher Kantonalbank
fabian.keller@zkb.ch

Urs Rhyner
Inventx Lab AG
urs.rhyner@inventx.ch

References

Alammar, J. (2019). *The Illustrated GPT-2 (Visualizing Transformer Language Models).* Retrieved from https://jalammar.github.io/illustrated-gpt2/

Alpaydin, E. (2020). *Introduction to machine learning.* MIT press.

Amel-Zadeh, A. (2017). *Why and How Investors Use ESG Information: Evidence from a Global Survey.* Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2925310

Ametepey, S. O., Aigbavboa, C., Thwala, W., & Addy, H. (2024). *The Impact of AI in Sustainable Development Goal Implementation: A Delphi Study.* Retrieved from https://www.mdpi.com/2071-1050/16/9/3858

Ankenbrand, T., Bieri, D., Reichmuth, L., Stengel, C., Wickihalder, S., & Yilmaz, E. A. (2023). *GPT for Financial Advice.* Retrieved from https://hub.hslu.ch/retailbanking/download/gpt-for-financial-advice/

Ankenbrand, T., Rhyner, U., & Yilmaz, E. (2023). *Quantum Computing and Artificial Intelligence in Finance.* Retrieved from https://www.researchgate.net/publication/376375411_Quantum_Computing_and_Artificial_Intelligence_in_Finance

Basu, K., Varanasi, S. C., Shakerin, F., Arias, J., & Gupta, G. (2021). *Knowledge-driven natural language understanding of english text and its applications.* In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 35, pp. 12554–12563).

Bildirici, Ozge Zeytin. (2023). *AI vs Machine Learning.* Retrieved from https://medium.com/illumination/ai-vs-machine-learning-fd642984e620

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., … others (2023). *Sparks of Artificial General Intelligence: Early Experiments with GPT-4. arXiv preprint arXiv:2303.12712.*

Busch, T., Baker, V., Brunner, K., Döbeli, S., Hess, K., Kimmerle, H., … Preussner, E. (2023). *Swiss Sustainable Investment Market Study 2023.* Swiss Sustainable Finance. Retrieved from https://marketstudy2023.sustainablefinance.ch/wp-content/uploads/2023/06/SSF_2023_MarketStudy.pdf

Cambria, E., & White, B. (2014). *Jumping NLP Curves: A Review of Natural Language Processing Research.* Retrieved from https://ieeexplore.ieee.org/document/6786458

Chen, X., Jia, S., & Xiang, Y. (2020). *A review: Knowledge reasoning over knowledge graph.* Retrieved from https://www.sciencedirect.com/science/article/pii/S0957417419306669

Ding, N., Qin, Y., Yang, G., Wei, F., Yang, Z., Su, Y., … Sun, M. (2023). *Parameter-efficient fine-tuning of large-scale pre-trained language models.* Retrieved from https://www.nature.com/articles/s42256-023-00626-4

Federal Data Protection and Information Commissioner. (2023). *Current data protection legislation is directly applicable to AI.* Retrieved from https://www.edoeb.admin.ch/edoeb/en/home/kurzmeldungen/2023/20231109_ki_dsg.html

Gartner. (2023). *What's New in Artificial Intelligence from the 2023 Gartner Hype Cycle.* Retrieved from https://www.gartner.com/en/articles/what-s-new-in-artificial-intelligence-from-the-2023-gartner-hype-cycle

Global Impact Investing Network. (2023). *2023 GIINsights.* Retrieved from https://thegiin.org/research/publication/2023-giinsight-series/

Grosan, C., & Abraham, A. (2011). *Intelligent systems* (Vol. 17). Springer.

Hassani, K., & Khasahmadi, A. H. (2020). *Contrastive Multi-View Representation Learning on Graphs.* Retrieved from https://arxiv.org/abs/2006.05582

Hinton, G., Vinyals, O., & Dean, J. (2015). *Distilling the Knowledge in a Neural Network.* Retrieved from https://arxiv.org/abs/1503.02531

Jeong, C. (2023). *A Study on the Implementation of Generative AI Services Using an Enterprise Data-Based LLM Application Architecture* (Vol. 03) (No. 04). Advances in Artificial Intelligence and Machine Learning. Retrieved from http://dx.doi.org/10.54364/AAIML.2023.1191

Ji, S., Pan, S., Cambria, E., Marttinen, P., & Yu, P. (2022). *A Survey on Knowledge Graphs: Representation, Acquisition, and Applications.* Retrieved from https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9416312

Khan, W., Daud, A., Khan, K., Muhammad, S., & Haq, R. (2023). *Exploring the frontiers of deep learning and natural language processing: A comprehensive overview of key challenges and emerging trends.* Retrieved from https://www.sciencedirect.com/science/article/pii/S2949719123000237?via % 3Dihub

Kharas, H., McArthur, J. W., & Ohno, I. (2019). *Leave No One Behind: Time for Specifics on the Sustainable Development Goals.* Brookings Institution Press.

Manakul, P., Liusie, A., & Gales, M. J. (2023). *SelfCheckGPT: Zero-Resource Black-Box Hallucination Detection for Generative Large Language Models. arXiv preprint arXiv:2303.08896*.

Mattman, B., & Stüttgen, M. (2023). *IFZ Sustainable Investments Studie 2023.* IFZ. Retrieved from https://zenodo.org/records/10281409

OpenAI. (online). *Models - OpenAI API.* Retrieved from https://platform.openai.com/docs/models/gpt-4-turbo-and-gpt-4

Ovadia, O., Brief, M., Mishaeli, M., & Elisha, O. (2023). *Fine-Tuning or Retrieval? Comparing Knowledge Injection in LLMs. arXiv preprint arXiv:2312.05934*.

Principles for Responsible Investment. (2018). *Impact Investing Market Map.* Retrieved from https://www.unpri.org/thematic-and-impact-investing/impact-investing-market-map/3537.article

Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: A Modern Approach.* Pearson.

Sachs, J., Schmidt-Traub, G., Mazzucato, M., Messner, D., Nakicenovic, N., & Rockström, J. (2019). *Six Transformations to achieve the Sustainable Development Goals.* Nature Sustainability. Retrieved from https://unstats.un.org/sdgs/report/2023/

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2020). *DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter.* Retrieved from https://arxiv.org/abs/1910.01108

SDSN Association. (online). *About Us.* Retrieved from https://www.unsdsn.org/about-us

Sejnowski, T. J. (2023). *Large language models and the reverse turing test. Neural Computation*, *35*(3), 309–342. Retrieved from https://direct.mit.edu/neco/article/35/3/309/114731/Large-Language-Models-and-the-Reverse-Turing-Test

Spiess-Knafl, W., & Scheck, B. (2017). *Impact Investing.* Palgrave Studies in Impact Finance. Retrieved from http://library.sadjad.ac.ir/opac/temp/19316.pdf

Swiss Bankers Association. (online). *Growth in sustainable finance continuing.* Retrieved from https://publications.swissbanking.ch/banking-barometer-2023/swiss-banking-outlook?overlay=Growth % 20in % 20sustainable % 20finance % 20continuing

Van Zanten, J. A., & Van Tulder, R. (2018). *Multinational enterprises and the Sustainable Development Goals: An institutional approach to corporate engagement.* Retrieved from https://link.springer.com/article/10.1057/s42214-018-0008-x

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., … Schmidt, D. C. (2023). *A prompt pattern catalog to enhance prompt engineering with chatgpt. arXiv preprint arXiv:2302.11382*.

Lucerne School of Business
Institute of Financial Services Zug IFZ
Campus Zug-Rotkreuz
Suurstoffi 1
6343 Rotkreuz

T +41 41 757 67 67
ifz@hslu.ch
hslu.ch/ifz

A study conducted by

HSLU Lucerne University of Applied Sciences and Arts

AACSB ACCREDITED

ISBN-Number
978-3-907379-30-1